DOCUMENT RESUME

ED 074 138 TM 002 501

AUTHOR	Kleinke, David J.
TITLE	An Index of Predictive Efficiency.
PUB DATE	Feb 73
NOTE	11p.; Paper presented at annual meeting of the American Educational Research Association (New Orleans, Louisiana, February 25-March 1, 1973)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Correlation; *Predictive Validity; Speeches; *Statistical Studies; *Testing
IDENTIFIERS	*Decision Theory Approach

ABSTRACT

An index that reflects the accuracy of selection associated with a predictive validity of "r" is presented. Based on Sheppard's thoerem on median dichotomies, it is a measure of improvement over chance assignment tc "accept" (or "reject"). Because the index is a measure of the accuracy of this assignment, rather than of variation from prediction throughout the distribution, the index is deemed to be a more appropriate measure than r squared or indices based on r squared when the purpose of testing is selection or placement. (Author/DB)

FILMED FROM BEST AVAILABLE COPY

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

ED 074138

TM 002 501

AN INDEX OF PREDICTIVE EFFICIENCY

David J. Kleinke

Syracuse University

Paper presented to meeting of American Educational Research Association

New Orleans, February 1973.

Abstract

A index that reflects the accuracy of selection associated with a predictive validity of $\underline{r}$ is presented. Based on Sheppard's theorem on median dichotomies, it is a measure of improvement over chance assignment to "accept" (or "reject"). Because the index is a measure of the accuracy of this assignment, rather than of variation from prediction throughout the distribution, the index is deemed to be a more appropriate measure than r^2 or indices based on r^2 when the purpose of testing is selection or placement.

Background

Merely knowing that the correlation between a set of test scores and score criterion measure is different from zero (at some level of statistical significance) is not enough for the evaluation of the coefficient. For a validity coefficient to be judged properly, one must "evaluate the size of the correlation in the light of the uses to be made of the test" (Anastasi, 1968, p. 130). Such evaluations appear to have traditionally taken one of two directions. The first will be deemed the "variance-accounted-for" approach, which has concentrated on r^2, the squared correlation between the predictor (test) and criterion (variable).[1] The second, called the "decision-theory" approach, is concerned with the accuracy of identification of positive and negative instances of criterion status such identification being on the basis of one's predictor status. The first has resulted in the derivation of a number of indices for interpreting correlation coefficients, which the second sees as generally being too "restrictive."

Variance-accounted-for

r^2, the "coefficient of determination" or "coefficient of association" is the proportion of variance on the criterion variable accounted for by variance on the predictor. It may be demonstrated algebraically that r^2 is the ratio of the variance of the predicted criterion scores to the variance of the obtained criterion scores. The complement of r^2, $k^2 = 1 - r^2$, has also been deemed to be useful. Known as the "coefficient of nondetermination" or "coefficient of nonassociation," it is obviously the proportion of variance on the criterion not accounted for by that on the predictor. Consideration of k^2 can lead to such statements as, "Since most correlations between tests of academic aptitude and college freshman grade point averages are in the order of .4, we can say that after sixty years of the mental measurement movement,

we have improved our technology to the point where only 85 percent of our prediction is error."

The square root of k^2 is of some interest also. Known as the "coefficient of alienation," it serves as the factor by which the standard deviation of the criterion is multiplied in order to produce the standard error of estimate. Of more immediate concern is the use of k in the computation of E, Hull's (1927) "index of forecasting efficiency," $E = \sqrt{1 - r^2}$ and is usually interpreted as representing the proportion of improvement over chance provided by a given validity coefficient.

The obvious logic of E and its conservative evaluation of various values of r have posed a dilemma for psychometricians for more than two score years. That E = .134 when r = .5 when verbalized as representing a "13.4 percent improvement over chance," is certainly irksome and possibly unreasonable. To produce an E of .5 requires that the criterion-related validity of a test be greater than .86, a value that could probably never be attained by any mental test.

Apologists have, for years, made attempts to reconcile this dilemma. As an example, consider Guilford's statement:

> Better tests, with validity coefficients of .60, have an E of about 20 per cent, and still better tests, when r = .75, have an E of about 34 percent. Although these efficiencies may also seem small, we must treat them in a relative, not an absolute sense. It is probable that the efficiency of predictions based upon the average unsystematic interview is less than 5 percent. With this as a base, the efficiency of tests looks much better. (1965, p. 378).

A more fruitful means of reconciling the conservatism of E with the demonstrable value of tests is by considering just what it is that is being improved over chance. What is being improved is the variation of obtained scores about the regression line estimating them.

Decision Theory

The decision-theoretic approach looks not at variation of obtained scores about a regression line but merely at correctness of classification. While there are many proponents of decision theory working in many areas of statistics, most are overlooked here to concentrate on those whose work applies most directly to the development of this paper. At the same time, workers whose contributions preceded the reification of decision theory as a discipline or at least a point of view are included.

The first important work to be considered is that of Taylor and Russell (1939). Their tables, showing the proportion of selected employees who prove to be successful under various combinations of selection ratios and success ratios, set the stage for modern psychometric decision theory, as set forth in Cronbach and Gleser (1965). Still, the Taylor-Russell tables are tables, not a relatively simple procedure yielding a single index.

The usefulness of a single index to evaluate a validity coefficient is fairly straightforward. One should be able to ask, "How much greater a coefficient do I need in order to have a measure that works twice as well as the measure I have at hand?"

Two approaches to the single-number problem here designated as "decision-theoretic" have been advanced. The first was that of Brogden (1946) who demonstrated that

> r is the ratio of the increase obtained by selecting above a given standard score on the predictor to the increase that would be obtained by selecting above the same standard score on the criterion itself (p, 68),

when the marginal distributions of the two variables are identical. Note, however, that the emphasis was on the scores obtained by those selected, not on the accuracy of the selection itself.

The second index with a decision-theoretic orientation is that of Jenkins (1953). His "index of selective efficiency," S, is equal to the ratio, (Successes - Failures)/(Successes + Failures), where "Successes" are "hits," in decision-theoretic terms. Jenkins did not report this, but if the splits are made at the medians of normally-distributed marginals or (equivalently) if there are dichotomous distributions with 50-50 splits then the tetrachoric r or the phi coefficient that would result would equal S.

The Index of Predictive Efficiency

Sheppard's theorem on median dichotomy (Kendall and Stuart, 1963) is that, for any value of r, the probability that a person who scores above the median on the predictor will also score above the median on the criterion is

$$P_2^0 = \frac{1}{4} + \frac{\text{arc sin } r}{2\pi} \quad , \qquad (1)$$

where arc sin r is expressed in radians, assuming that the marginals are distributed normally and the joint distribution is bivariate normal.

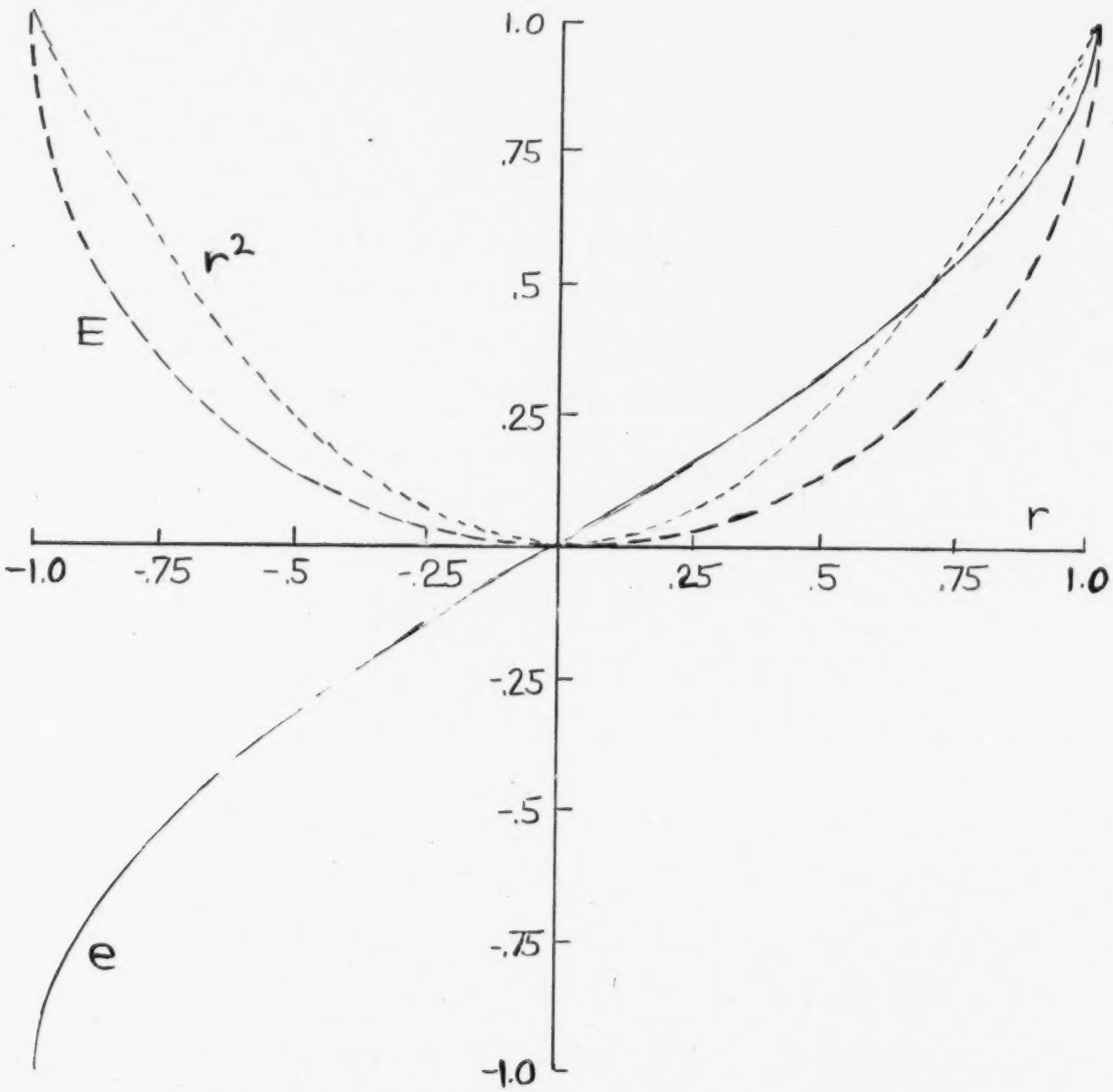
When $r = 0$, $P_2^0 = .25$; when $r = 1$, $P_2^0 = .5$. (arc sin r)/2π thus varies from 0 to .25 and represents the proportion of individuals for whom selection (or placement) is improved over chance selection by making use of a predictor of validity r. This value, when divided by .25, the proportion for whom prediction could be improved, yields

$$e = \frac{2 \text{ arc sin } r}{\pi} \qquad (2)$$

This index is called the "index of predictive efficiency." Its curve, together with those of r^2 and E, is plotted in Figure 1.

Examination of Figure 1 reveals certain interesting features of e. First, it is equal to r at $r = 0$ and $r = \pm 1$. Further, unlike functions

Figure 1. r^2, E, and e for $-1 \leq r \leq 1$.



of r^2, it is negative when r is negative. Its absolute value is, except at $r = \pm 1$ or $r = 0$, always greater than that of E. Its absolute value is greater than that of r^2 from $r = -.707$ to $r = .707$. It must be noted here that arc sin $.707 = 45^\circ$. $e = r^2 = k^2$ at that point, where the vectors representing the predictor and criterion are at 45°. That angle represents a correlation halfway between zero and unity. Finally, between $r = -.707$ and $r = .707$, the curve of e is approximately linear.

An informal empirical check of e was conducted, using one hundred "examinees" and four variables. The variables were selected at random to have first essentially zero correlation and then progressively higher ones, approaching unity. The original form of the marginals was approximately rectangular. They then underwent transformations to both approximate normality and marked negative skewness. The obtained values of r ranged from -.03 to .97.

The results were frustrating. Somehow, although the "unrelated" variables had r's of .03, .01, and -.03, for the rectangular, normal, and skewed cases, respectively, they had observed e's of approximately .2 for the .25, .5, and .75 slection ratios. However, the plot of the visual centroids of the various values of the observed e's followed that of Equation (2). Moreover, there appeared to be no systematic relationships among magnitude of r, shape of marginal, and e.

It must be concluded that e needs further study. Its properties are provocative. As an instance, that it is approximately linear in the range of validity coefficients that are reasonable for mental tests sheds new light on the Brodgen and Jenkins approaches, so long in the dark. The most obvious direction of this further study is an extensive exploration of

empirically-derived values of e, with varied marginal distributions, bivariate distributions, and selection ratios.

FOOTNOTE

[1] The use of "predictor" and "criterion" in this paper are for convenience and are not intended to limit the generality of the principles contained herein.

REFERENCES

Anastasi, A. Psychological testing (3rd Ed.). New York: Macmillan, 1968.

Brogden, H. E. On the interpretation of the correlation coefficient as a measure of predictive efficiency. Journal of Educational Psychology, 1946, 37, 65-76.

Cronbach, L. J. and Gleser, G. C. Psychological tests and personnel decisions. Urbana, Ill.: University of Illinois Press, 1965.

Guilford, J. P. Fundamental statistics in psychology and education. (4th Ed.). New York: McGraw-Hill, 1965.

Hull, C. L. The correlation coefficient and its prognostic significance. Journal of Educational Research, 1927, 15, 327-338.

Jenkins, W. L. An index of selective efficiency (S) for evaluating a selection plan. Journal of Applied Psychology, 1953, 37, 78.

Kendall, M. G. and Stuart, A. The advanced theory of statistics, V. I, Distribution Theory. New York: Hafner, 1963.

Taylor, H. C. and Russell, J. T. The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. Journal of Applied Psychology, 1939, 23, 565-578.